

## ABSTRACT OF THE DISCLOSURE

Techniques for determining user types based on multi-modal clustering are provided. The topology, content and usage of a document collection or web site is determined. The user paths are identified using longest repeating subsequence techniques and a multi-modal information need vector is determined for each significant user path. Multi-modal vectors for each document in the significant path, content, uniform resource locators, inlink and outlink multi-modal vectors are determined and combined based on path position and access frequency. Multi-modal clustering is performed based on a multi-modal similarity function and a specified measure of similarity using a type of multi-modal clustering such as K-means or wavefront clustering. The identified clusters may be further analyzed based on changes to the weighting of the corresponding content, url, inlinks and outlinks multi-modal feature vectors.

2025 RELEASE UNDER E.O. 14176